MBMG
Metabarcoding & Metagenomics

Metabarcoding and Metagenomics 7: 273–304 (2023)
DOI: 10.3897/mbmg.7.109085

Research Article

Metagenomic exploration of endosymbionts and pathogens in the tropical lineage of *Rhipicephalus sanguineus* sensu lato (s.l.) ticks in Colombia

Luisa Paez-Triana[1], Giovanny Herrera[1], Laura Vega[1], Diego García-Corredor[1,2], Martin Orlando Pulido Medellín[2], Alberto Paniz-Mondolfi[3], Marina Muñoz[1], Juan David Ramírez[1,3]

1 *Centro de Investigaciones en Microbiología y Biotecnología-UR (CIMBIUR), Facultad de Ciencias Naturales, Universidad del Rosario, Bogotá, Colombia*

2 *Grupo de Investigación en Medicina Veterinaria y Zootecnia GIDIMEVETZ, Universidad Pedagógica y Tecnológica de Colombia, Tunja, Colombia*

3 *Molecular Microbiology Laboratory, Department of Pathology, Molecular and Cell-based Medicine, Icahn School of Medicine at Mount Sinai, New York, USA*

Corresponding author: Juan David Ramírez (juand.ramirez@urosario.edu.co; juan.ramirezgonzalez@mssm.edu)

Abstract

Rhipicephalus sanguineus sensu lato (s.l.), commonly known as the domestic dog tick, is a globally distributed tick. This species plays a significant role in human and animal health, as well as in economy due to its ability to infest livestock. The use of shotgun metagenomics has proven valuable in identifying tick-borne pathogens and key members of the tick microbiome, including endosymbionts. However, the application of shotgun metagenomics in *R. sanguineus* s.l. ticks in Latin America remains largely unexplored. Therefore, our objective aimed to explore and further analyze the metagenome of the tropical lineage of *R. sanguineus* s.l. ticks in Colombia. Through our analysis, we identified the three most prevalent pathogens harbored by these ticks, namely: *Anaplasma phagocytophilum*, *Francisella tularensis*, and *Theileria equi*. However, the most abundant microorganism detected was not a pathogen but the endosymbiont *Coxiella mudrowiae*. Interestingly, *Coxiella mudrowiae* exhibited significant negative correlations with several pathogens and other endosymbionts. Furthermore, we successfully constructed 27 medium-quality metagenome-assembled genomes (MAGs) for this microorganism, enabling us to conduct a pangenome analysis by comparing them with available genomes and identifying proteins of interest, such as those involved in vitamin B synthesis. This study represents the first implementation of shotgun metagenomics as a methodology to expand our understanding of pathogens and endosymbionts in the circulating tropical lineage of *R. sanguineus* s.l. ticks in Colombia. The findings of this research serve as a foundation for the development of prevention and mitigation strategies against pathogens transmitted by this tick species. Information gained from this study can contribute to the improvement of public health measures and veterinary practices aimed at controlling the impact of tick-borne diseases.

Key words: Metagenome, microbiome, *R. sanguineus* s.l, shotgun-metagenomics, tick

OPEN ACCESS

Academic editor: Alexander Probst
Received: 5 July 2023
Accepted: 2 October 2023
Published: 20 November 2023

Citation: Paez-Triana L, Herrera G, Vega L, Garcia-Corredor D, Pulido Medellín MO, Paniz-Mondolfi A, Muñoz M, Ramirez JD (2023) Metagenomic exploration of endosymbionts and pathogens in the tropical lineage of *Rhipicephalus sanguineus* sensu lato (s.l.) ticks in Colombia. Metabarcoding and Metagenomics 7: e109085. https://doi.org/10.3897/mbmg.7.109085

Copyright: © Luisa Paez-Triana et al. This is an open access article distributed under terms of the Creative Commons Attribution License (Attribution 4.0 International – CC BY 4.0).

Introduction

Ticks are blood-feeding arthropods that serve as obligate vectors, capable of transmitting a wide range of pathogens, surpassing other types of vectors in

terms of diversity (Liu and Bonnet 2014). Among the hard-bodied ticks belonging to the family Ixodidae, several species have significant implications for public and animal health (Venzal et al. 2003). One such species is *Rhipicephalus sanguineus* sensu lato (s.l.), commonly known as the brown dog tick, which has a global distribution (Dantas-Torres 2010). This tick species displays endophilic and monotropic behavior, primarily infesting dogs, but it has also been found infesting wildlife, horses, and humans in rare instances (Dantas-Torres et al. 2013). *R. sanguineus* s.l. holds great importance in both human and animal health, exerting direct and indirect effects on various hosts, including economically significant livestock (Dantas-Torres et al. 2013). Within *R. sanguineus* s.l., five different lineages have been identified using mitochondrial molecular markers, namely *R. sanguineus* s.l. (tropical lineage), R. sp. I, R. sp. II (temperate lineage), R. sp. III, and R. sp. IV (Dantas-Torres et al. 2013). Among these lineages, *R. sanguineus* s.l. and R. sp. II are considered the primary lineages, with *R. sanguineus* s.l. being predominantly distributed in tropical regions, while R. sp. II is commonly found in temperate zones. However, it should be noted that certain countries exhibit a co-circulation of both lineages. Currently, only the circulation of the tropical lineage has been documented in Colombia (Paez-Triana et al. 2021).

It is widely recognized that ticks have the ability to harbor and transmit a diverse range of microorganisms, making them important vectors for diseases (Walker et al. 2000). While ticks can acquire numerous pathogens during their feeding, not all of them are capable of reproducing and being transmitted to new hosts. However, *Rhipicephalus sanguineus* s.l. is known to be a competent transmitter of a wide array of pathogens, including bacteria, protozoa, fungi, nematodes, and viruses (Jia et al. 2020). These tick-borne diseases (TBDs) contribute significantly to the overall burden of vector-borne diseases (VBDs), impacting human and animal health as well as national economies, especially in developing regions (Couper and Swei 2018; Ergunay et al. 2022). Therefore, there is a pressing need for comprehensive surveillance to identify pathogens and assess the potential risks to public health (Woolhouse and Gowtage-Sequeria 2005; Ergunay et al. 2022). However, studying the ecological community of ticks involves more than just identifying pathogens. The tick microbiota also includes commensal and endosymbiotic organisms (Pollet et al. 2020), understanding endosymbionts as those organisms that provide a benefit to the tick. These benefits include enhancing reproductive fitness, facilitating nutritional adaptation, influencing development and reproduction, providing defense against environmental stressors, and modulating tick immunity (Pollet et al. 2020). Also, an impact of this community on pathogen transmission has been observed. For example, in *Dermacentor andersoni*, negative correlations were found between parasites and endosymbionts like *Anaplasma marginale* and *R. belli* and positive between *Francisella* endosymbionts and *F. novicida* (Gall et al. 2016).The mechanisms involve the secretion of compounds that affect pathogen growth, either positively or negatively or the competition between the two organisms for nutrients, leading to reduced pathogen growth, or even the endosymbiont's impact on the tick's immune system results in increased pathogen growth (Gall et al. 2016). Many of these have been identified as endosymbionts based on their genomic characteristics (Gottlieb et al. 2015), the absence of virulence factors indicating their potential non-pathogenic nature,

the presence of essential metabolic pathways for tick survival (Buysse and Duron 2021), and even through in-vivo experiments demonstrating the significant benefits of their presence in ticks (Ben-Yosef et al. 2020). Among these endosymbionts are the *Coxiella*-like endosymbionts (CLE), *Wolbachia* endosymbionts (WE), *Rickettsia* endosymbionts (RE), and *Francisella* endosymbionts (FE), among others (Hussain et al. 2022). Despite research on the potential pathogenicity of some of these organisms, it is currently assumed that they do not cause pathogenicity in humans and is limited to specific animals (Hussain et al. 2022). However, their significance in ticks as endosymbionts has indeed been widely recognized, rendering them a community of great importance for study.

Moreover, comprehensive understanding of the tick microbiota is crucial for unraveling the complex interactions within this community and their implications for tick biology and disease transmission. Traditional methods used to identify microorganisms in ticks have limitations in terms of sensitivity, as they only capture a fraction of the tick's microbiota (Wade 2002; Batool et al. 2021). These methods primarily focus on the most common pathogens associated with tick-borne diseases. They involve techniques such as culturing, microscopy, and PCR coupled with Sanger sequencing (Batool et al. 2021). In contrast, metagenomic techniques, particularly shotgun metagenomics, offer a more comprehensive understanding of the tick's microbial composition by identifying all genes and genomes present (Xia et al. 2015). This approach allows for the assessment of microbial diversity, abundance, and the analysis of complete or partial genomes (Xia et al. 2015). By utilizing metagenomics, we can strategically survey tick-borne pathogens, evaluate the associated risks, and gain insights into the mechanisms of disease transmission by these microorganisms. Additionally, this approach enables the identification of other ecologically significant communities, such as endosymbionts. Metagenomics goes beyond taxonomic classification and provides initial insights into the functional roles performed by these microorganisms (Bonnet and Pollet 2021), considering the diverse pathways of interaction between these communities and ticks. Previous studies have effectively employed this methodology. For example, a study conducted in Palestine focused on *Rhipicephalus* spp. and *Haemaphysalis* spp., successfully identified various pathogenic species and endosymbionts. They generated genomes for *Rickettsia massiliae*, Candidatus *Rickettsia barbariae*, and *Coxiella* endosymbionts (Ravi et al. 2019). Similarly, investigations involving *Amblyomma* species have revealed important pathogens such as *Coxiella burnetii* and *Francisella tularensis* (Ergunay et al. 2022). In the case of *R. sanguineus* s.l., several studies have explored its microbiota using amplicon-based sequencing targeting the 16S rRNA gene. Pollet et al. identified Proteobacteria as the dominant phylum, with the genus *Rickettsia* and the family Coxiellaceae being the most abundant in these ticks (Pollet et al. 2020). Another study by René-Martellet et al. confirmed the prevalence of the genus *Rickettsia* and observed a high abundance of *Coxiella* and *Bacillus*, particularly in the tropical lineage, which exhibited the highest *Coxiella* abundance (René-Martellet et al. 2017). Additionally, Cassia-Luzzi et al. compared the microbiota composition among different lineages of *R. sanguineus* s.l. and between embryos and female adults, revealing a significant abundance of Proteobacteria in all lineages and Actinobacteria and Firmicutes in the tropical lineage (Luzzi et al. 2021).

The study of *R. sanguineus* s.l. ticks in South America, particularly in Colombia, is limited despite its significance in developing and tropical countries. A study by Cotes-Perdomo et al. (2020) conducted PCR analysis and identified *Anaplasma marginale*, *Babesia vogeli*, *Babesia bigemina*, and *Coxiella* spp. (*C. mudrowiae* and endosymbionts) in ticks collected from dogs in the Magdalena department. However, the lack of comprehensive research hinders our understanding of the potential impact of this tick species on domestic animals and humans in the country. It is crucial to consider the potential transmission of medically and veterinary important diseases by *R. sanguineus* s.l., including Rocky Mountain spotted fever, human granulocytic anaplasmosis, Lyme disease, and canine babesiosis. Despite the significant advantages of shotgun metagenomics in identifying a wide range of organisms, this technology has not yet been applied to *R. sanguineus* s.l. ticks in Latin America. However, this approach could provide valuable insights into the interactions and functions of the entire microbiome of these ticks, considering the diverse interaction pathways between these communities and the tick. It has been observed that pathogens can activate mechanisms and manipulate tick responses to facilitate their infection, while ticks can limit pathogen infection, and high bacterial infections can enhance tick survival. Based on the aforementioned information, our objective is to analyze the metagenome of the tropical lineage of *R. sanguineus* s.l. ticks in the Santander and Casanare departments in Colombia. By employing shotgun metagenomics, we aim to enhance our understanding of the microbial composition and potential pathogens associated with these ticks, thereby contributing to the development of effective prevention and control strategies for tick-borne diseases in the region.

Methods

Sample collection and genetic material extraction for sequencing

A total of 38 adult ticks, collected from dogs, were obtained from the Casanare and Santander departments in Colombia. In the municipality of Yopal, Casanare (5°19'50"N, 72°23'26"W), we collected nine females and eleven males. Similarly, in the municipality of Puente Nacional, Santander (5°52'38"N, 73°40'43"W), we collected ten males and eight females (Suppl. materials 1, 6). The ticks were manually removed from lesions that were clearly caused by ticks using tweezers and preserved in Zymo Shield solution (DNA/RNA Shield, Zymo Research, Cat.: R1100). Morphological identification of the ticks as *R. sanguineus* s.l. (tropical lineage) was done with the assistance of taxonomic keys. For DNA extraction, a partial fraction of each tick was used, always the half of the tick divided longitudinally. The ticks were washed three times in PBS to remove impurities. Subsequently, the DNeasy Blood & Tissue Kit (Qiagen) was utilized following the manufacturer's instructions, except for extending the incubation time of the disruption buffer by 12 hours at 56 °C. The concentration of each extracted DNA sample was evaluated using the Qubit 2.0 Fluorometer and the dsDNA High Sensitivity kit (Thermo Fisher) following the recommended protocol. The quality of the DNA was assessed by electrophoresis on 1.5% agarose gels. To perform the sequencing, the samples were sent to Novogene (Sacramento, CA, USA) and sequenced using the Illumina platform for shotgun metagenomics.

The sequencing was carried out on the NovaSeq 6000 platform, employing paired-end 150 bp reads. Each sample generated approximately 6 Gbytes of storage for the raw data per sample. The number of reads is specified in Suppl. material 6.

Read-based bioinformatic analysis

A graphical summary for all the methods is in Suppl. material 2. The quality assessment of the raw reads was conducted using FastQC v.0.11.9 (Andrews 2010), and MultiQC v.1.6 (Ewels et al. 2016) was employed for summarizing the results. Subsequently, Trimmomatic v.0.38 (Bolger et al. 2014) was utilized to filter the reads based on the following criteria: MINLEN:150, AVGQUAL:20, and TRAILING:20. The "Iluminaclip" parameter was applied to eliminate adapter sequences from the reads using the TruSeq3-PE sequence. On average, < 0.00% of reads were removed in this filtering step. The filtered reads, which exhibited high quality, were then aligned to the available genome of *R. sanguineus* s.l (Jia et al. 2020), which can be found in the GenBank with the accession number GCA_013339695.1. This alignment was performed using Bowtie2 v.2.4.4 (Langmead and Salzberg 2012). For taxonomic assignment of the remaining reads, Centrifuge v.1.0.3-beta (Kim et al. 2016) was employed, utilizing the program's default database. The resulting outputs were transformed to Kraken-Report format using the Centrifuge-kreport function. The visual representation of the data was accomplished using Pavian (Breitwieser and Salzberg 2020) and the ggplot2 package in RStudio (Wickham 2008).

Identification of pathogens and endosymbionts

To identify pathogens and endosymbionts, a custom database (Number of genomes (N): 2153, Number of genomes remove (NR): 524) was created by compiling relevant information from the literature. Genomes of *Anaplasma* (N: 50, NR: 2), *Rickettsia* (N: 146, NR: 73), *Francisella* (N: 1075, NR: 161), *Ehrlichia* (N: 44, NR: 2), *Borrelia* (Lyme Disease) (N: 379, NR: 220), *Coxiella* (N: 158, NR: 0), *Babesia* (N: 5, NR: 2), and *Theileria* (N: 4, NR: 3) available in NCBI (RefSeq Annotated) were downloaded (Available genomes until July 2023). The database was then refined by excluding species that are not transmitted by ticks and those that are not known to circulate in the Americas. For the identification of endosymbionts, genomes of the *Wolbachia* (N: 292, NR: 62) were included in the genomes of endosymbionts belonging to the genus *Rickettsia* and *Coxiella*. Only the genera identified as endosymbionts of insects or arachnids were retained. The information of the genomes used are in Suppl. material 7.The created database was used with Centrifuge v.1.0.3-beta (Kim et al. 2016) to match against the reads. Centrifuge employed a minimum length of partial hits (--min-hitlen) of 95 and a k of 1 (k classification parameter). The visual representation was made using the same specifications as mentioned earlier.

Analysis of correlation and abundance

To explore differences in read abundance between sexes and departments, Wilcoxon tests were performed. The significance level was set at 0.05, and the

analysis was conducted using RStudio software. This statistical test helped identify any significant variations in read abundance, providing valuable insights into potential differences associated with sex and geographical location. Additionally, to examine the relationships within and between different ecological communities, such as endosymbionts and pathogens, the Spearman's non-parametric rank-order correlation with the Benjamini-Hochberg correction was employed. This statistical analysis was conducted using the "psych" package in RStudio. Significant correlations were identified when the p-value was less than 0.05. Strong correlations were defined as having a correlation coefficient (ρ) below -0.75 or above 0.75. The resulting correlations were then visualized using Cytoscape 3.9.0 (Shannon et al. 2003; Gustavsen et al. 2019). Visualization was facilitated by utilizing R packages such as "igraph," "ggraph," and "Rcy3." These tools aided in creating a network representation of the correlations, allowing for a comprehensive understanding of the relationships among the various ecological communities.

Virulence factors and resistance markers

To identify virulence factors, the Basic Local Alignment Search Tool (BLAST) was utilized. The host-filtered reads, obtained after aligning the reads to the Virulence Factors Database (VFDB), were subjected to BLASTn analysis, considering a minimum of 95% identity and at least an e-value of 7.04E-07 as a match. This approach allowed for the identification of genetic elements associated with the virulence of microorganisms (Altschul et al. 1990; Liu et al. 2019). In addition, to detect antibiotic resistance markers, the Resistance Gene Identifier (RGI) tool was employed. The clean reads obtained from the sequencing data were used as input for RGI. The CARD (Comprehensive Antibiotic Resistance Database) version 3.1.3, released on 5 July 2021, was employed as the reference database for identifying antibiotic resistance genes (Alcock et al. 2020).

Bioinformatic analysis for the construction of metagenomes (MAGs)

The clean reads were subjected to assembly using MetaSpades v3.15.3 (Bankevich et al. 2012; Nurk et al. 2017). Binning for each sample was performed using MetaBAT (Kang et al. 2019), CONCOCT (Alneberg et al. 2014), and MaxBin (Wu et al. 2014; Wu et al. 2016), with subsequent refinement using DAS Tool (Sieber et al. 2018), all employing default parameters. The quality of the assembled genomes was assessed using CheckM v1.1.3 (Parks et al. 2015; Bowers et al. 2017), following established quality parameters for metagenomic genomes (MAGs). Taxonomic assignment of the high-quality bins was carried out using the Genome Taxonomy Database (GTDB-Tk) v1.7.0 (Chaumeil et al. 2020), employing default parameters. Only bins with completeness greater than 50% and contamination less than 10% (medium-quality MAGs) were obtained. All bins were taxonomically classified as Candidatus *Coxiella mudrowiae* and selected for further analysis. The obtained MAGs were annotated using the PROKKA program (Seemann 2014) using default parameters and visualized using PROKSEE (Stothard et al. 2018). All MAGs taxonomy assignments were double checked with pubMLST (Jolley et al. 2018).

Pangenome and comparative genomics of MAGs

A step of dereplication was perform for all the MAGs, using the program deRep (Olm et al. 2017), The parameter "genomeInfo" was included, containing the quality of the MAGs obtained from CheckM. Additionally, the "ignoreGenomeQuality" command was used due to the average quality of the MAGs. However, to perform the comparative genomics analysis and investigate the genetic diversity and variability of the species, we use the 27 obtained MAGs for *C. mudrowiae* with available ones and conducted a pangenome analysis. The genomes of Ca. *Coxiella mudrowiae* were downloaded from NCBI and combined with the herein recovered, Ca. *Coxiella mudrowiae* RSt (GCF_001077715.1) and Ca. *Coxiella mudrowiae* RSA-CAT (GCF_002804145.1). The Panaroo tool (Tonkin-Hill et al. 2020) was employed to analyze these genomes, considering a 95% identity threshold and their presence in at least 95% of the compared genomes. The parameters used included "--clean-mode strict -a core --aligner mafft --core_threshold 0.95". The output from Panaroo was utilized to visualize phylogenetic trees using iTol (Interactive Tree Of Life) (Letunic and Bork 2019), and Phandango (Hadfield et al. 2018) was used to create graphical representations of the genomes. However, because the analysis conducted using all obtained genomes seemed to produce biased results, we carried out an additional analysis. In this subsequent analysis, we excluded four specific MAGs while following the same protocol as previously described. To determine the relationships between the genomes obtained in this study and other Ca. *Coxiella mudrowiae* genomes, a phylogenetic network was constructed using the Neighbor Joining method in the SplitsTree5 program (Huson and Bryant 2006).

Furthermore, a comparative genomics analysis was conducted. Genomes identified as endosymbionts of ticks from the *Coxiella* genus and *Coxiella burnetii* RSA (GCF_000007765.2) were downloaded from NCBI. Amino acid trees of 40 proteins were constructed using the Uprot program in Taxxo (https://github.com/giraola/taxxo/wiki/Taxxo-wiki), which was executed within RStudio. These 40 proteins are universal phylogenetic markers that are single-copy and have been successfully used in various studies. ANIb (Average Nucleotide Identity using Blast) values were calculated between the complete genomes using the Taxxo tool. Genome pairs with an ANI score higher than 95.0% were considered to belong to the same species (Jain et al. 2018). Lastly, a distance analysis of SNPs was performed between the different genomes using snp-dist v. 2.4.1 and snp-sites v. 2.4.1 (Page et al. 2016), with default settings.

Functional analysis of MAGs

The annotation outputs from PROKKA were utilized to identify Clusters of Orthologous Groups (COG) with eggNOG-mapper v2 (Cantalapiedra et al. 2021), employing default settings. Additionally, the presence of crucial metabolic pathways was verified using KEGG KAS pathways (Kanehisa et al. 2016) with default setting for metagenomes. Specifically, genes associated with the biosynthesis of vitamin B and cofactors were selected and extracted from our MAGs and *Coxiella burnetii* RSA. The Ugene program (Okonechnikov et al. 2012) was employed to calculate the percentage of similarity between these genes. Furthermore, the corresponding genes from other *Coxiella* endosymbiont

genomes were included and aligned using the MAFFT program, using default settings. Subsequently, phylogenetic trees were constructed using IQ-TREE (Minh et al. 2020) with an ultrafast bootstrap of 1000 repetitions. The best replacement model was chosen by jModelTest (Posada 2008). Finally, the trees were visualized using iTol (Minh et al. 2020).

Results

Read-based bioinformatic analysis

A total of 18.2 to 38.1 Mb of data per sample was obtained. However, the reads related to the host represented 88.5% of the reads (SD: ± 6). After the filter, only a small percentage of the sequences were successfully classified, with a mean of 19% (SD: ± 5). This left an average of 81% of the sequences unclassified in each sample. Among the classified sequences (Suppl. materials 8, 9), bacteria accounted for the majority (3% to 24%), followed by archaea (0.07% to 0.09%), and viruses (0.02% to 0.04%). The family Coxiellaceae was found to be predominant in all samples (Fig. 1A), with a *Coxiella* endosymbiont representing a significant proportion ranging from 48% to 95% of the classified reads in bacteria. Apart from this species, the bacterial community in ticks was dominated by genera such as *Bacillus*, *Clostridium*, *Pseudomonas*, *Rickettsia*, *Streptomyces*, *Propionibacterium*, *Burkholderia*, *Fusobacterium*, *Lactobacillus*, and *Campylobacter*, in descending order of abundance. In terms of archaea (Fig. 1B), the dominant families were Methanobacteriaceae, Methanocaldococcaceae, Methanosarcinaceae, Thermoplasmataceae, and Sulfolobaceae. Each family was represented by a corresponding genus: *Methanothermobacter*, *Methanocaldococcus*, *Methanosarcina*, *Thermoplasma*, and *Sulfolobus*, respectively. As for viruses, they were distributed among the families Baculoviridae, Myoviridae, and Siphoviridae. However, the majority of reads were attributed specifically to the genera *Alphabaculovirus* and *Betabaculovirus*, indicating a clear dominance of the Baculoviridae family. Although other viral families and genera were identified, the number of reads for each of them was minimal, making it challenging to accurately determine their taxonomic assignment.

Although the overall composition remained consistent across all departments and sexes, significant differences were found in the most abundant genera (Suppl. material 3). Within the bacteria (Suppl. material 3: fig. S3A) *Clostridium*, *Fusobacterium*, and *Rickettsia* showed greater abundance in Santander. These same genera exhibited higher abundance in females within Casanare. On the other hand, *Corynebacterium* and *Propionibacterium* demonstrated higher abundance in Casanare. Additionally, along with these, *Staphylococcus* and *Streptomyces* displayed higher abundance in males within Santander. Furthermore, *Propionibacterium* also showed higher abundance in Casanare. Regarding archaea (Suppl. material 3: fig. S3B), *Methanobrevibacter*, *Methanococcus*, *Methanosarcina*, and *Sulfolobus* exhibited higher abundance in Casanare. Meanwhile, *Methanocaldococcus*, *Methanohalophilus*, and *Thermolasma* showed greater abundance in Santander. Only *Thermococcus* displayed sex-specific differences, with higher abundance in males.

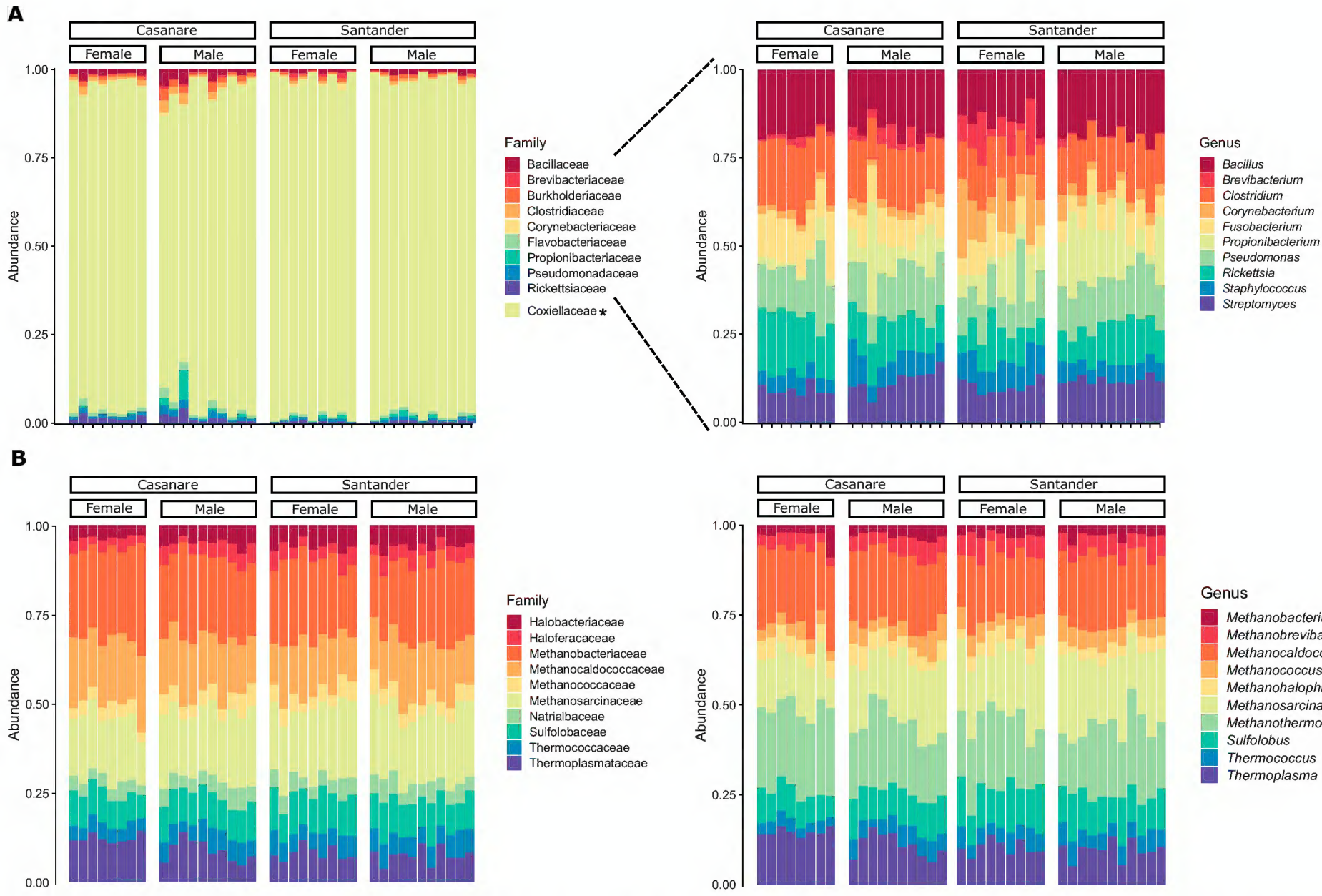

Figure 1. Relative abundance of the top 10 **A** bacteria **B** archaea classified reads per tick, categorized by department and sexes. Families are indicated on the left side of the corresponding panel, while genera are presented on the right. Coxiellaceae *: For the bacterial genus (right panel), we excluded the reads from the Coxiellaceae family as they represent most of the relative abundance of the genus and were all assigned to *Coxiella* endosymbionts, which hindered the assessment of the abundance of other genera.

Identification of pathogens and endosymbionts in the samples

Regarding pathogens (Suppl. material 11), the analysis revealed the dominance of three species in the samples (Fig. 2A). *Francisella tularensis* was the most prevalent in all ticks (MEAN: 23%, SD: 8), followed by *Anaplasma phagocytophilum* (MEAN: 7%, SD: 1), and to a lesser extent, *Theileria equi* (MEAN: 7%, SD: 3). Additional pathogens from the genus *Babesia*, *Ehrlichia*, *Borrelia* and *Coxiella* were also identified on the top 10, albeit with lower abundance ranging from 2% to 0.02% (Fig. 2A). Other pathogens were detected in the samples, but their abundance was significantly lower (Suppl. material 4: fig. S4A), with percentages below 0.02% and a very limited number of reads per species. On the other hand, among the identified endosymbiotic species (Suppl. material 11), *Coxiella mudrowiae* accounted for a substantial proportion of the reads, with a mean of 47% (SD: 18) of the reads assigned with this database. Without *C. mudrowiae*, the top 10 endosymbionts identified were (Fig. 2B): WE of *Atemnus politus*, *Corcyra cephalonica*, *Culex molestus*, *Bemisia tabaci*, and *Nilaparvata lugens*, CLE of *Dermacentor marginatus*, *Rhipicephalus microplus*, *Ornithodoros amblus* and *Amblyomma americanum*, and RE of *Ixodes scapularis* (Fig. 2B). These endosymbionts exhibited abundance ranging from 0.02% to 3%. Other endosymbionts were identified but with even lower relative abundances (Suppl. material 4: fig. S4B), with percentages below 0.02%, and a minimal number of reads per species.

Analysis of correlation and abundance

Although no noticeable differences between sex and location were observed in the heatmaps (Fig. 2A, B), statistical analysis using Wilcox test (Suppl. material 12) revealed significant variations in the relative abundances of the top three pathogens with the highest number of reads (Fig. 2C). These differences were characterized by higher abundances in the Santander department, which was also consistent for the majority of the top 10 identified pathogens (Suppl. material 4: fig. S4C, Suppl. material 12). Moreover, significant differences in pathogen abundance were found between sexes within the Casanare department in *F. tularensis* and *T. equi* pathogens, where males consistently exhibited higher pathogen abundance (Fig. 2C). A similar pattern was observed for these two pathogens within the Santander department, although not statistically significant. This sex-based pattern was consistent across most of the species among the top 10 pathogens (Suppl. material 4: fig. S4C). However, *A. phagocytophilum* exhibits an inverse pattern in the Santander department. In this region, females displayed a significantly higher relative abundance (Fig. 2C). Regarding the endosymbionts, *Coxiella mudrowiae* and most other species showed significant differences between departments. Only *Coxiella mudrowiae* exhibited higher abundance in the Casanare department, while the rest of the endosymbionts were more abundant in Santander (Fig. 2D, Suppl. material 4: fig. S4D). Furthermore, a significant difference was observed between sexes within the Casanare department (Fig. 2D). *Coxiella mudrowiae* displayed higher abundance in females, whereas the remaining endosymbionts exhibited greater abundance in males (Fig. 2D). Taking into account the

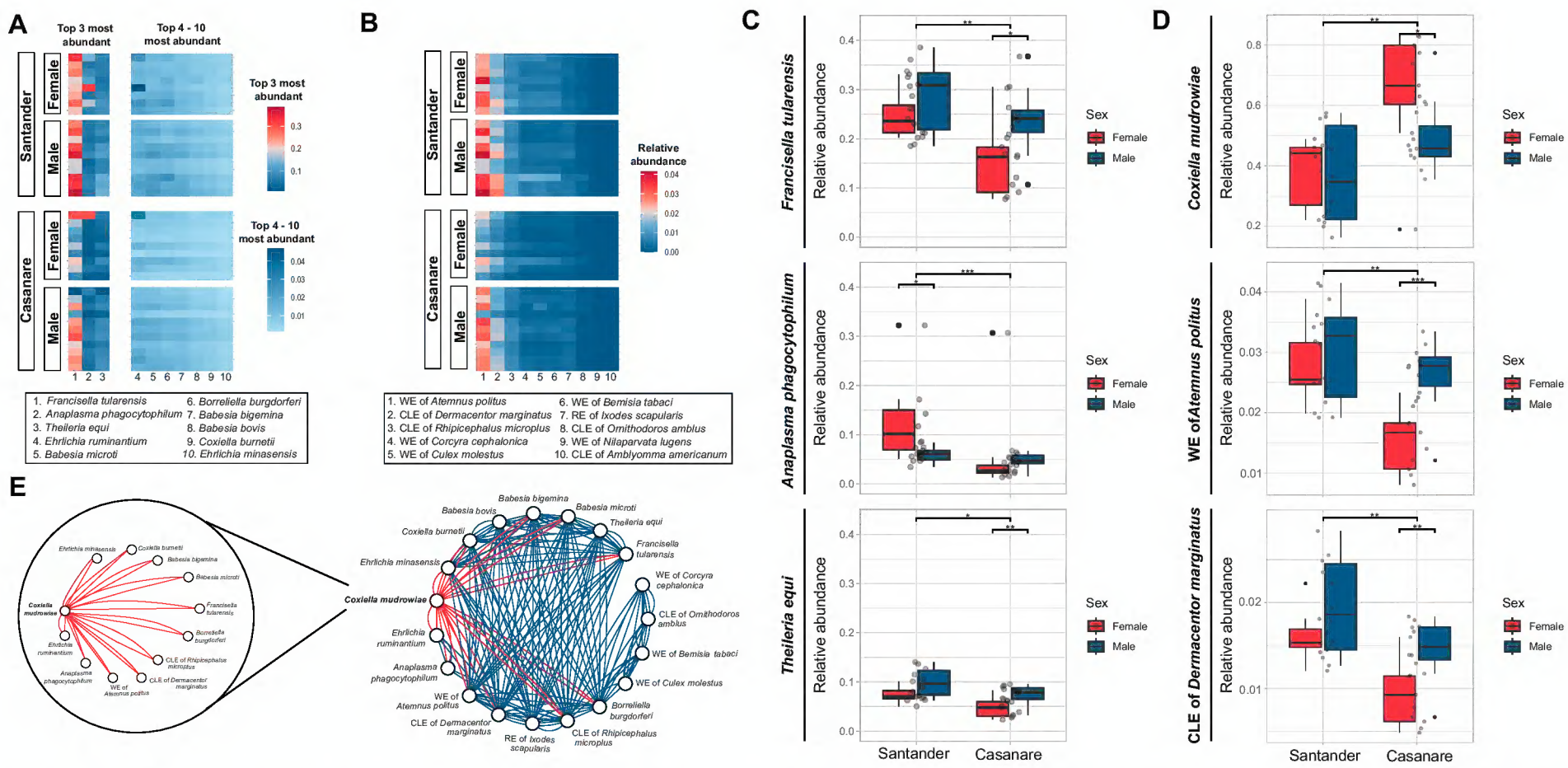

Figure 2. Pathogens and endosymbionts present in each tick. **A** top 10 pathogens **B** top 10 Endosymbionts (excluding *C. mudrowiae*) **C** differences in relative abundance between departments and sexes for the top 3 pathogens **D** differences in relative abundance between departments and sexes for the top 3 endosymbionts **E** networks of positive (blue) and negative (red) correlations among the identified species. The significances are ranked as follows: The significances are ranked as follows: *: < 0.05 **: < 0.01 ***: < 0.001. WE: *Wolbachia* endosymbiont. RE: *Rickettsia* endosymbiont. CLE: *Coxiella* like endosymbiont.

differences in abundance, a correlation test was conducted among the top 10 species from each ecological community, revealing strong positive correlations between endosymbionts and pathogens, as well as within the same ecological communities (Fig. 2E). However, negative correlations were found between *Coxiella mudrowiae* and eleven different species (three endosymbionts and eight pathogens).

Virulence factors and resistance markers

The ticks showed a low abundance of reads (>52 reads per sample) related to virulence factors (Suppl. material 13). The majority of these reads were associated with the virulence factors Hemin binding protein b (hbpB), 6-phosphogluconate dehydrogenase (gnd), and adhesion and penetration protein (app). These factors are typically associated with *Bartonella quintana*, *Klebsiella pneumoniae* and *Neisseria meningitidis*, respectively. In contrast, a significantly higher number of reads were observed for the resistance markers. Five markers accounted for approximately 80% to 90% of the relative abundance of all the resistance markers. (Suppl. material 14). These markers were identified as APH(3")-Ib, kdpE, aadA24, sul1, and acrS, listed in order of their abundance. Notably, the kdpE marker exhibited a significantly higher abundance (22%) in a male sample from Casanare compared to the remaining samples, where its abundance remained below 10%. These markers are associated with pathogens such as *Escherichia coli*, *Klebsiella pneumoniae* and *Salmonella enterica*.

Bioinformatic analysis for the construction of metagenomes (MAGs)

A total of 27 medium-quality MAGs belonging to the species Ca. *Coxiella mudrowiae* were successfully assembled (Fig. 3A). This assignment was validated using pubMLST (Jolley et al. 2018). The completeness of the assemblies ranged from 87.21% to 89.44%, with contamination ranging from 0.58% to 1.2% (Suppl. material 15). Our quality analysis revealed the absence of 20 genes that are typically used by checkM, along with the duplication of one of these genes (Suppl. material 6). We carried out a dereplication step. Upon observing the primary clustering dendrogram, despite identifying two separate clusters, both exhibited a MASH ANI identity percentage greater than 90%, thereby consolidating into a single cluster (Suppl. material 5: fig. S5A). This unified cluster was then analyzed in the secondary clustering dendrogram, where the pair-wise distance between all organisms was minimal (Suppl. material 5: fig. S5A). However, considering the distinction among the 4 genomes in the primary dendrogram, we made the decision to conduct independent comparative genomics and pangenome analysis for each of the 27 obtained MAGs, as well as for the available genomes. Pangenome analysis revealed that these four MAGs contained a significant number of genes in the accessory genome, which were absent in the other MAGs and available genomes (Fig. 3B, Suppl. material 5: fig. S5C). As a result, these four MAGs exhibited significantly different genome lengths compared to the others (Suppl. materials 5, 15). This led to an imbalance in both gene percentages and quantities between the core genome and the accessory genome, posing challenges in effectively confining the structure of the genome in our MAGS. As an illustration, one of these longer

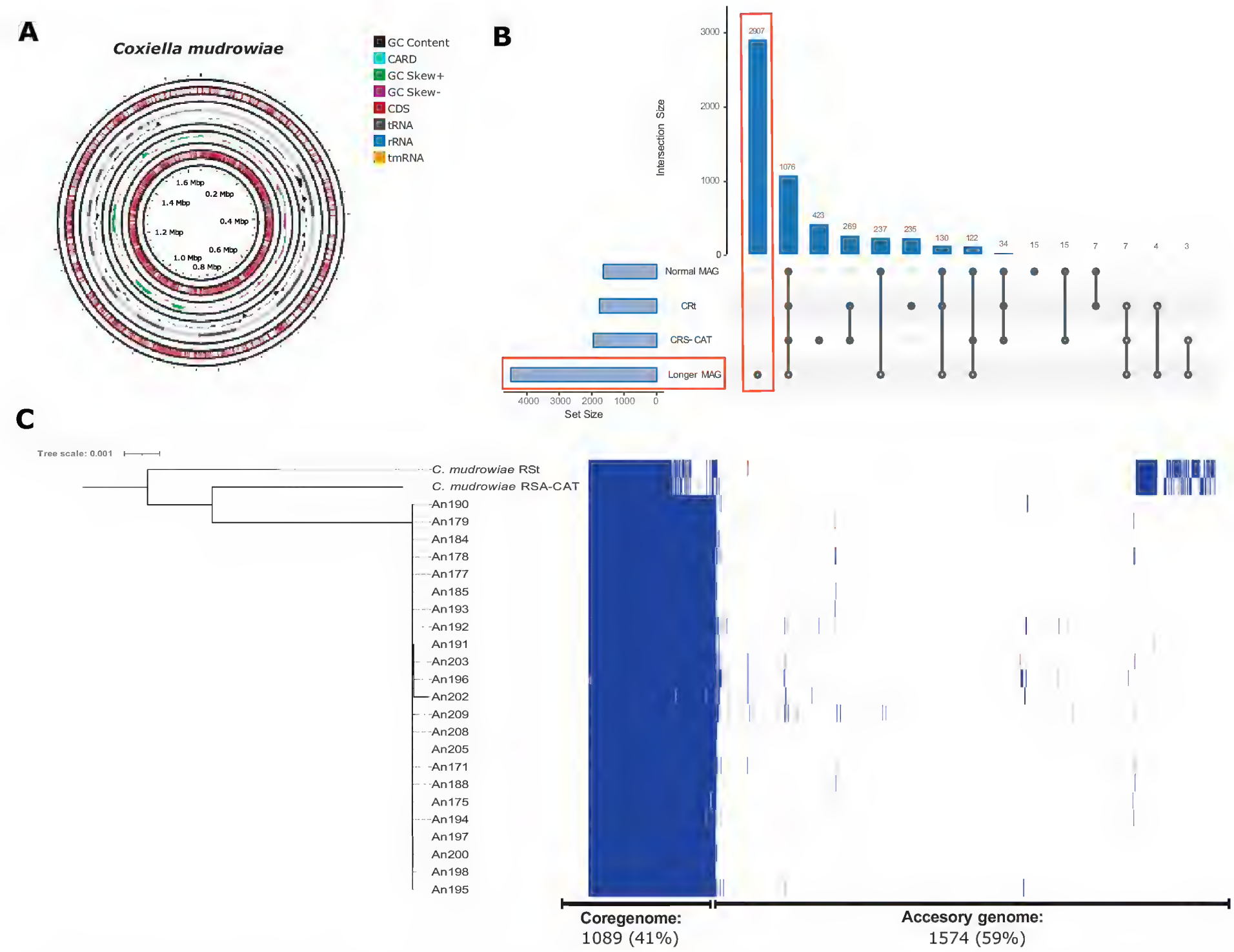

Figure 3. Assembled MAGs belonging to *C. mudrowiae* species. **A** example (Sample An171) of the assembled MAGs and its components **B** shared genes between a representative of normal length MAG, a representative of longer length MAG, *C. mudrowiae* RSt, and *C. mudrowiae* RSA-CAT. This is a representation utilizing a reference MAG for each of the groups (Longer and Normal length). The red boxes highlight the Longer MAG, its contained gene count, and the substantial number of unique genes associated with it. **C.** Evolutionary insights between samples (without the Longer MAGs) based on the core genome (left panel) of the MAGs and available genomes, compared with a matrix where the core and accessory genes were either present or absent, which was graphically represented (right panel).

MAGs encompassed a total of 2907 unique genes in contrast to one normal length MAG and available genomes (Fig. 3B). In comparison, a representative of a Normal length MAGs demonstrated the presence of 252 genes exclusively found within them, distinct from the genomes available in the dataset (Fig. 3B). Notably, these extended-length MAGs predominantly featured hypothetical proteins, with no discernible plasmids identified among their genetic content. To accurately elucidate the composition of both the core genome and accessory genome, an additional pangenome analysis was executed (Fig. 3C), wherein these aforementioned extensive genomes were excluded from the analysis (this procedure was made with the same protocol detailed in the methods section). This procedure revealed a core genome comprising 41% (1089 genes) and an accessory genome accounting for 59% (1574 genes). To assess the coding potential of our MAGs compared to the available genomes of *C. mudrowiae* and *C. burnetii*, we examined the COG (Clusters of Orthologous Groups)

sets (Suppl. material 16). We observed a similarity in the number of genes across all groups between the available genomes of *C. mudrowiae* and our MAGs, although there were slight reductions in certain groups. Additionally, a decrease in gene numbers compared to *C. burnetii* was observed in groups such as Transcription, Replication, recombination and repair, Cell cycle control, cell division, chromosome partitioning, among others.

The phylogenetic analysis based on the core genome (Fig. 3C) exhibited a clustering pattern where the MAGs from our study formed a distinct group, separated to some extent from the two available genomes. This separation was further supported by the SNP and genome distance analyses (Fig. 4A, B), as the MAGs from our study exhibited greater genetic distance from the available genomes compared to the genetic distance observed among the available

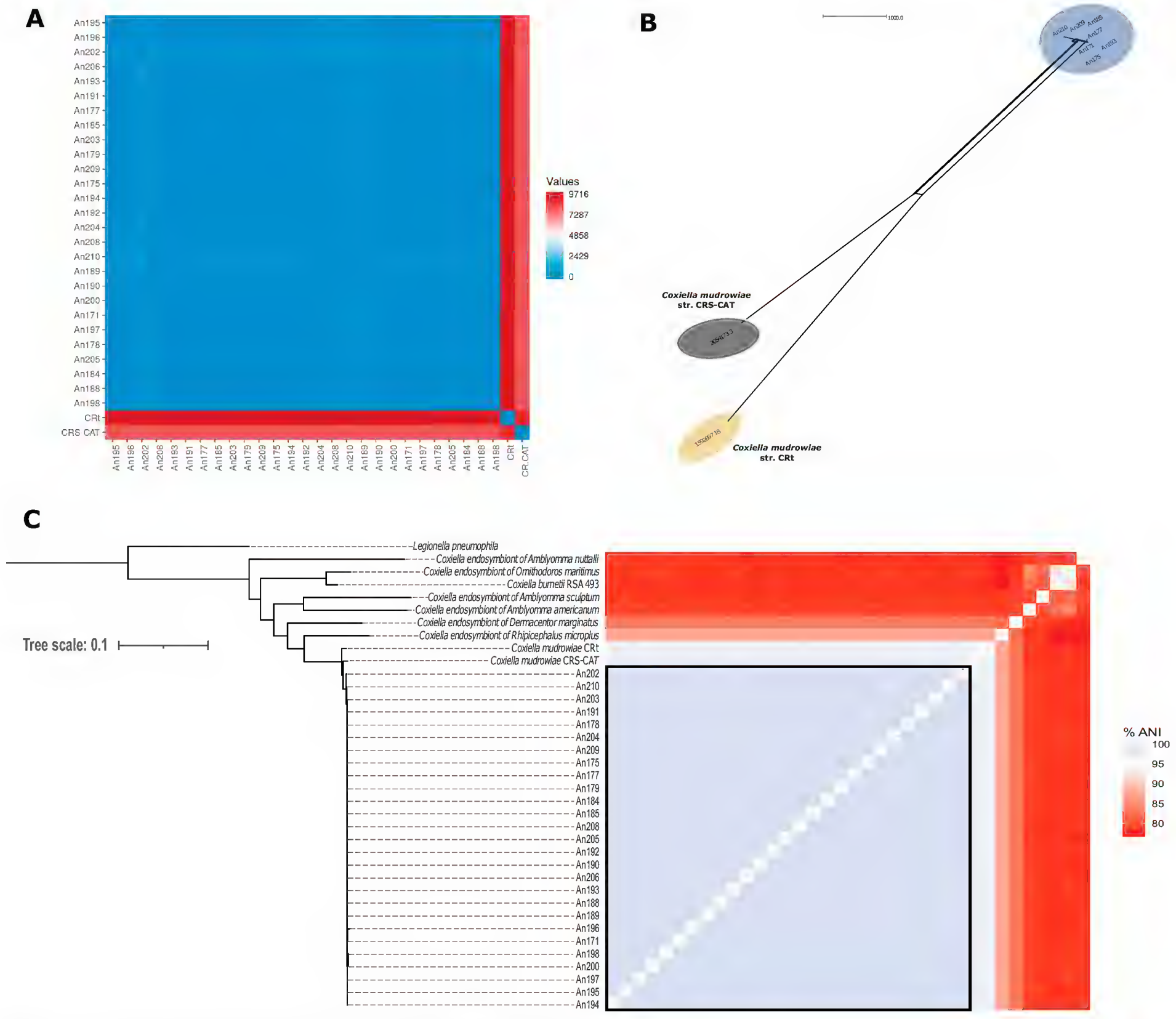

Figure 4. Comparative genomics and phylogenetic analysis of the obtained MAGs in relation to those available for *C. mudrowiae*. **A** genetic distance of SNPs between the available genomes and the obtained MAGs **B** phylogenetic network representing the relationships among the MAGs and available genomes **C** phylogenetic tree (bootstrap 1000 replicates, and jModelTest was used to select the best substitution method) based on 40 highly conserved proteins across various CLE, *Coxiella burnetii*, and MAGs (left panel). This phylogenetic tree is complemented by an ANIb analysis (right panel) for the entire genome conducted on CLE, *Coxiella burnetii*, and MAGs, with a threshold set at 95%.

genomes themselves. Considering these findings, a comprehensive phylogenetic tree was constructed using 40 proteins, incorporating the available genomes of *Coxiella mudrowiae*, the available *Coxiella* endosymbiont genomes, and our MAGs. This analysis, combined with ANIb calculations on these genomes, revealed a clear clustering pattern between the available genomes of *Coxiella mudrowiae* and our MAGs. Additionally, the ANIb values were over 95%, indicating no species differentiation (Jain et al. 2018) (Fig. 4C). This analysis provided compelling evidence for the distinction between *Coxiella mudrowiae* (represented by our MAGs and available genomes) and the other endosymbionts, as well as *Coxiella burnetii*.

Previous studies have suggested that the putative role of this endosymbiont involves synthesizing essential nutrients that are not readily available in the blood, considering the obligate hematophagy of this vector (Gottlieb et al. 2015; Brenner et al. 2021). Building upon this knowledge and the information available on the genes responsible for the synthesis of vitamin B and cofactors in the existing genomes of *C. mudrowiae*, we investigated the presence of these genes in our MAGs. Our analysis revealed the presence of complete pathways for the synthesis of Biotin (B7), Riboflavin (B2), and the cofactors CoA and FAD in our MAGs (Fig. 4A). However, we did not identify the genes phoA, panE, and panD, which are involved in the biosynthesis pathways of Pyridoxine (B6) and Folic Acid (B9). These pathways have been previously reported in both *Coxiella* endosymbionts (CLE) and *Coxiella burnetii*. Nevertheless, upon comparing the proteins encoded by these genes, we observed very low similarity between our MAGs and *Coxiella burnetii* (Fig. 5A). This suggests significant non-synonymous mutations in these genes. Phylogenetic trees further support this observation, as they distinctly separate *Coxiella burnetii* from *Coxiella mudrowiae* and most previously reported endosymbionts (Figs 5B, C).

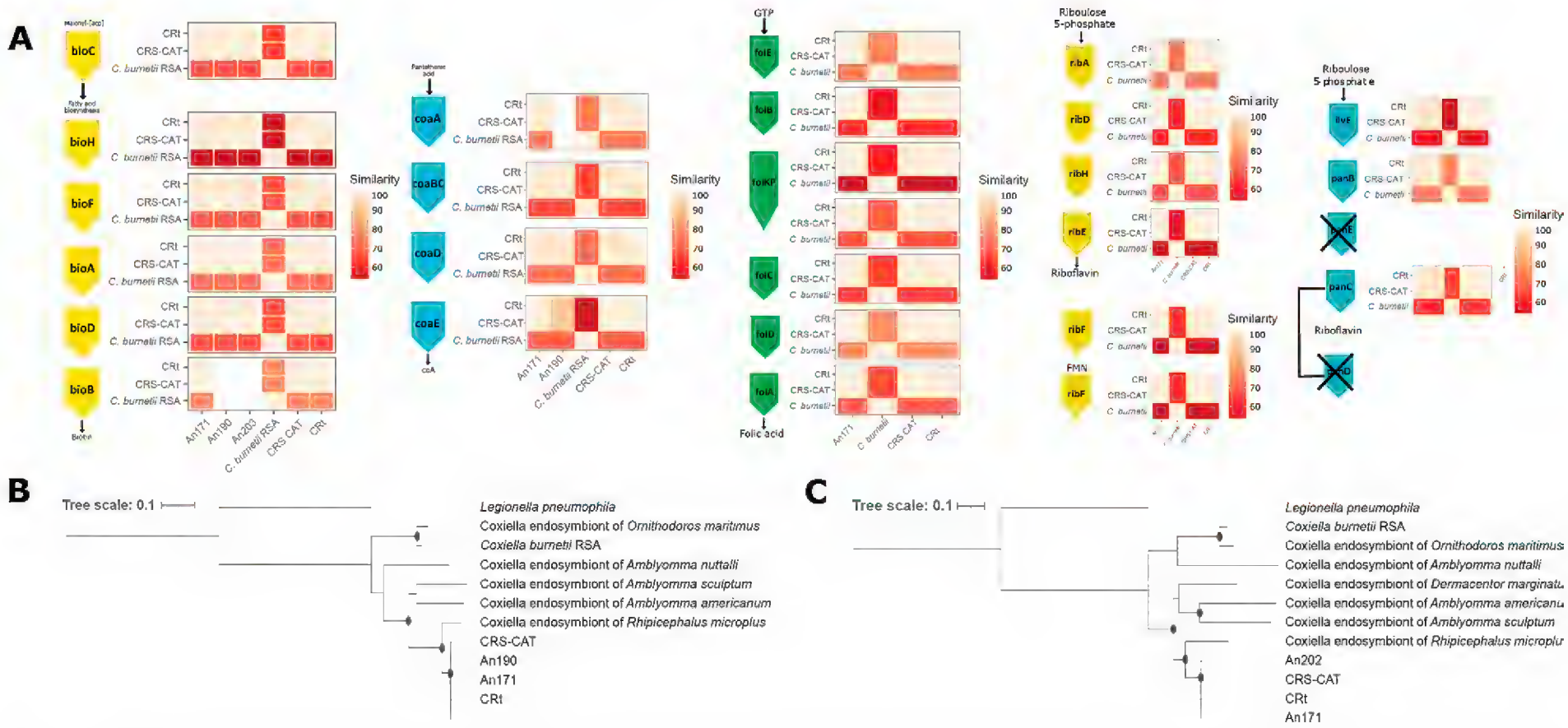

Figure 5. A genes involved in A. Vitamin B synthesis and B. cofactors identified in MAGs, *C. mudrowiae* RSt, *C. mudrowiae* RSA-CAT, and *C. burnetii*, along with the identity percentages of the proteins produced by them **B** phylogenetic tree (Bootstrap 1000 replicates) of BioB protein of all avaible genomes for CLE, *C. burnetii* and MAGs **C** phylogenetic tree (Bootstrap 1000 replicates) of coaE protein of all available genomes for CLE, *C. burnetii* and MAGs.

Discussion

Ticks are recognized as the second most important vectors of diseases after mosquitoes (Chen et al. 2023). The Center for Disease Control and Prevention (CDC) reports that tick-borne diseases (TBDs) accounted for over 75% of the 650,000 reported vector-borne diseases in the United States between 2004 and 2016 (Rosenberg et al. 2018). The advent of advanced sequencing technologies, such as shotgun metagenomics, has revolutionized disease research by enabling rapid and accurate identification of emerging and re-emerging pathogens (Chen et al. 2023). This approach has also shed light on the diverse ecological communities that exist within ticks, providing a comprehensive understanding of tick microbiomes (Xia et al. 2015). Such comprehensive knowledge is crucial for the prevention and identification of both veterinary and human diseases, offering valuable insights into these disease-carrying vectors. Metagenomics has the potential to drive innovative strategies in vector control and the prevention of TBDs. By leveraging the information obtained through metagenomic studies, we can develop targeted interventions and enhance our ability to identify and combat TBDs effectively. This approach holds great promise for improving the health outcomes of both animals and humans by strengthening our prevention and identification efforts in the field of tick-borne diseases.

Here we employed shotgun metagenomics to characterize the microbiome of 38 *R. sanguineus* s.l. specimens. The predominant bacterial composition in our samples was found to be from the Coxiellaceae family (Fig. 1A), which is consistent with previous studies on the tropical lineage of *R. sanguineus* s.l. ticks worldwide (René-Martellet et al. 2017; Portillo et al. 2019; Luzzi et al. 2021). Notably, research by Luzzi et al. (2021) demonstrated a significant similarity in bacterial composition between embryos and adults, suggesting the potential for transovarial transmission of *Coxiella* spp. within this family. The identification of endosymbionts within the Coxiellaceae family (Fig. 1A) further emphasizes their importance in *R. sanguineus* s.l. ticks, as they have been consistently observed across different continents and studies. In addition to the Coxiellaceae family, our study identified other genera such as *Pseudomonas*, *Rickettsia*, *Propionibacterium*, and *Bacillus*, which have been previously associated with this tick species (Lalzar et al. 2012; René-Martellet et al. 2017). Furthermore, the presence of genera like *Clostridium*, *Streptomyces*, *Burkholderia*, *Fusobacterium*, *Lactobacillus*, and *Campylobacter*, although not previously reported in *R. sanguineus* s.l., has been documented in other tick species (Karim et al. 2017; Papa et al. 2020). It is likely that these bacteria may be acquired from the environment or host, such as the skin or cavities. It is important to note that certain bacteria, such as *Bacillus* and *Rickettsia* species, have been reported to exhibit a higher prevalence in one of the sexes (René-Martellet et al. 2017); our study found significant differences in various genera between sexes and departments. Including *Rickettsia*, where females displayed higher abundance, this observation aligns with the findings of René-Martellet et al. Despite these differences, the fundamental composition of the top 10 genera remains consistent across departments and sexes.

The presence of archaea and viruses in the samples was minimal in terms of abundance, but a diverse range of genera was identified within these groups (Fig. 1B). Archaea in ticks have received limited attention and research due to

the scarcity of available data. The genus *Methanobacterium* has been previously detected in other tick species like *Ixodes granulatus* (Chen et al. 2023). As far as we know our study provides the first identification of archaea in this tick species using metagenomics. However, it is important to note that their role in ticks remains unclear and further investigation is necessary to understand their potential interactions and contribution to the tick's bacterial and eukaryotic communities. With this finding it seems that the methane can play a role in the tick. However, a more comprehensive approach is required to explore the involvement of archaea in tick physiology and health and unravel their significance in tick interactions. In addition, this study identified the presence of the Baculoviridae family, which consists of arthropod viruses (Herniou and Jehle 2007). Specifically, the genera *Alphabaculovirus* and *Betabaculovirus*, known as lepidopteran-specific nucleopolyhedroviruses (NPVs) (Herniou and Jehle 2007), were found to be dominant among the DNA viruses detected in the samples. The presence of these viruses suggests their potential involvement in tick biology and their interaction with the tick's natural environment. Further research is needed to elucidate the specific roles and implications.

To identify the pathogens and endosymbionts present in the samples and explore their interactions, a database was created using available genomes. As a result, three major pathogens were detected in the samples: *A. phagocytophilum*, which causes human granulocytic anaplasmosis (HGA) and anaplasmosis in animals (Karshima et al. 2022); *F. tularensis*, the causative agent of tularemia (Telford and Goethert 2020); and *T. equi*, the causative agent of equine piroplasmosis (PE) (Wise et al. 2013). The vector competence of *R. sanguineus* s.l for these pathogens is currently unknown. However, previous studies have reported the detection of pathogen DNA in *R. sanguineus* s.l specimens worldwide (Ghafar and Amer 2012; Zhang et al. 2012; Lopes de Carvalho et al. 2016; Cabrera et al. 2022; Rocafort-Ferrer et al. 2022). Moreover, research conducted in Colombia has also identified these pathogens in both hosts and ticks, suggesting their likely circulation within the country (Máttar and Parra 2006; Molina-Guzmán et al. 2019; Santodomingo et al. 2019; Bonilla-Aldana et al. 2022; Cabrera et al. 2022). Considering the significance of these pathogens in terms of animal and human health, including the potential for humans to act as accidental hosts for these diseases (Cabrera et al. 2022), and the potential economic impact on a developing nation (Santodomingo et al. 2019), it is crucial to further investigate the vector competence of *R. sanguineus* s.l for these pathogens. However, it is important to consider in the case of *F. tularensis* that, as of now, genomes of FE have not been identified in RefSeq. Moreover, it has been observed that metagenomic studies can lead to confusion in distinguishing endosymbionts from their pathogenic species (Buysse and Duron 2021). Therefore, future research should be directed towards confirming the presence of *F. tularensis* or FE in these ticks. Even so, long-term surveillance studies focusing on these three tick-borne diseases should be implemented in Colombia. It is important to note that there is a hypothesis suggesting that the reported cases of TBDs in humans within the country may be underestimated due to underreporting (Máttar and Parra 2006). Additionally, considering the tendency of this tick species to feed on domestic animals such as dogs, cattle, and horses, there is a significant likelihood of human-animal contact in Colombia, highlighting the need for further research in this area.

On the other hand, a significant abundance of reads corresponded to various endosymbionts (Fig. 2B). It is well-known that ticks harbor a diverse microbiome primarily composed of non-pathogenic organisms, and studying their ecology and evolutionary dynamics is crucial. In particular, a substantial number of reads aligned with *Coxiella* endosymbionts. However, some reads were also classified as *Wolbachia* endosymbionts (WE). Recent studies have reported the presence of *Wolbachia* in *R. sanguineus* s.l. ticks (Chao et al. 2021). Nevertheless, it is essential to identify and further investigate these endosymbionts due to the potential implications of *Wolbachia* in controlling vector-borne diseases in other arthropods (Chao et al. 2021). Therefore, future research on this tick species should expand, encompassing genome sequencing, characterization of their role as endosymbionts, and even exploring their potential experimental use in vector control. Regarding the identified *Coxiella* spp., a clear dominance of *Coxiella mudrowiae* (Fig. 2D) was observed, a candidate species previously identified and described as endosymbiont. *C. mudrowiae* has been detected in various geographical locations and different tick species (Ravi et al. 2019; Rahal et al. 2020). This candidate was recognized as an ongoing endosymbiont based on several genomic features that are characteristic of endosymbionts. These include a notable reduction in genome size when compared to *C. burnetii*, a pronounced rate of pseudogenization, and the pseudogenization of characteristic virulence genes such as the Dot/Icm Type IVB secretion system (T4BSS), a crucial system for the pathogenicity of *C. burnetii* (Gottlieb et al. 2015; Tsementzi et al. 2018; Brenner et al. 2021). However, imaging data or clinical evidence to verify whether this organism could potentially become a pathobiont were not found. Consequently, further development of this research area is imperative, given its widespread prevalence worldwide. Nevertheless, since all current evidence suggests its role as an endosymbiont and no pathogenic traits have been identified, it will be classified as an endosymbiont throughout the paper. Additionally, multiple *Coxiella*-like endosymbionts (CLE) were detected within the same tick individuals, indicating that multiple CLEs can establish themselves within the tick microbiota. This phenomenon has been previously discovered through the analysis of the 16S rRNA gene in tick species, including *R. sanguineus* s.l. (Machado-Ferreira et al. 2016). It is plausible that each endosymbiont provides a different function to the tick and coexists or that they compete for colonization. However, we observed a negative correlation between *C. mudrowiae* and the other CLE (Fig. 2E), suggesting that competition for colonization is more likely. Nonetheless, conducting further comprehensive research on both species is crucial to better understand their interactions, whether it be WE-WE, WE-CLE, or CLE-CLE, as well as to comprehend the possible coevolution occurring between different endosymbionts and ticks, and even among the endosymbionts themselves.

It is well-established that endosymbionts engage in ongoing interactions with the transmitted pathogens, which can either facilitate, restrict, or impede their transmission (Bonnet et al. 2017). Examining the correlations, *Coxiella mudrowiae* showed negative associations with multiple pathogens and endosymbionts. This finding was further supported by observing regions with a higher abundance of this CLE, where pathogen read abundance was reduced. Similarly, negative correlations were identified with other endosymbionts. In the Casanare department, where *Coxiella mudrowiae* was more prevalent, the

remaining endosymbionts showed lower abundance, and vice versa (Fig. 2C, D). Previous studies have highlighted the importance of investigating these interactions. For example, in *Dermacentor andersoni* ticks, a negative correlation was found between pathogens such as *Anaplasma marginale* and *Rickettsia belli* and endosymbionts, while a positive correlation was observed between *Francisella* endosymbionts and *Francisella novicida* (Gall et al. 2016). Therefore, further exploration of these negative interactions involving *Coxiella mudrowiae*, which is known to be abundant in this and other tick species, is crucial. This investigation should encompass not only next-generation sequencing techniques but also biological studies aimed at validating and enhancing our understanding of these interactions.

Similarly, considering the prominence of this endosymbiont in our samples, we assembled multiple MAGs and compared them to the genomes constructed thus far. Currently, two genomes have been characterized, one derived from *R. sanguineus* s.l and another from *R. turanicus* (Gottlieb et al. 2015; Tsementzi et al. 2018), both originating from Israel. These genomes highlight the functional role and significance of *Coxiella* endosymbionts in ticks, as they are involved in vitamin synthesis, cofactor production, and other metabolic processes (Tsementzi et al. 2018). In our comparison of the MAGs to these reference genomes, we observed the absence of several genes (Fig. 3B, C), which could be attributed to their incomplete nature. However, analysis of the clusters of orthologous groups (COGs) revealed that the quantities were highly similar to those in the available genomes, albeit with a reduction in certain functional groups compared to *C. burnetii*. This reduction has been previously documented and is associated with pseudogenization and gene loss resulting from the obligate endosymbiotic nature of this bacterium (Gottlieb et al. 2015).

Furthermore, the MAGs formed a distinct monophyletic group separate from the assembled genomes. However, when analyzing average nucleotide identity (ANI) percentages and constructing a phylogenetic tree based on 40 informative proteins, the MAGs clustered with known *Coxiella* species and indicated a low degree of diversity within the country. This aligns with previous research that has attributed the limited intrapopulational diversity of *C. mudrowiae* to the restricted number of hosts, as represented by the two *Coxiella* strains derived from different tick species (Gottlieb et al. 2015). These findings are consistent with our observations and the anticipated limited diversity within the country (Paez-Triana et al. 2021). However, the substantial single-nucleotide polymorphism (SNP) differences between our MAGs and the available genomes could be attributed to the considerable geographic distance separating the reported origin countries of these genomes. Therefore, we hypothesize that a distinct strain of *C. mudrowiae* is circulating within Colombia. Such an undertaking would enable comprehensive genomic and phylogenetic analyses, thereby enhancing our understanding of the co-evolutionary mechanisms between *C. mudrowiae* and its tick host.

Finally, we conducted an assessment of specific genes that have been previously reported in this endosymbiont (Fig. 4), particularly those involved in vitamin B synthesis (Tsementzi et al. 2018; Brenner et al. 2021). We observed the presence of these genes, consistent with findings from the other two genomes of *C. mudrowiae*. However, we noted the absence of panD and panE, which have been documented in other *Coxiella*-like endosymbionts, such as the CLE of

Rhipicephalus sanguineus s.l I (Brenner et al. 2021). Nevertheless, it has been postulated that these missing genes may be substituted by other genes present in the genome of *C. mudrowiae* (Tsementzi et al. 2018). Additionally, previous studies have indicated that these genes are also present in *C. burnetii*, suggesting their essential role in the pathways required by both organisms and their interaction with their respective hosts (human and tick) (Brenner et al. 2021). However, upon analyzing the protein sequences, we discovered substantial variations that resulted in low similarity between the proteins and even divergence in the protein trees (Fig. 4). Despite previous investigations in proteomics highlighting the significance of these genes and proteins in the endosymbiotic association between *C. mudrowiae* and its tick host (Cibichakravarthy et al. 2022), it is crucial to assess how these dissimilarities in the proteins might impact their structure and functionality compared to what *C. burnetii* can produce. Understanding the metabolic aspects of this endosymbiont could potentially contribute to future biotechnological strategies for controlling tick populations and managing tick-borne diseases both within the country and globally. Previous studies have indicated that the absence of *Coxiella*-like endosymbionts in ticks results in delayed post-feeding development, prolonged feeding periods, and reduced fertility and fecundity. Larvae lacking CLE endosymbionts have been observed to exhibit unsuccessful feeding (Ben-Yosef et al. 2020). Overall, further investigations focusing on the genetic and functional characteristics of these genes and proteins are necessary to fully comprehend their roles in the endosymbiotic relationship between *C. mudrowiae* and ticks. Such understanding can inform the development of innovative approaches for tick control and the management of tick-borne diseases.

Despite certain limitations, such as competition for genomic resources between ticks and their pathogens, which can limit the number of assigned reads, and challenges associated with taxonomic assignment due to the scarcity of research on vector microbiomes, particularly in ticks, understanding the microbiome of these vectors is of utmost importance. It not only allows us to identify the pathogens that the tropical lineage of *R. sanguineus* s.l. could transmit or carry to a wide range of animals, including humans due to their close contact with domestic dogs, but it is also essential for the control of diseases transmitted by these ticks. This is particularly crucial considering the increasing emergence of resistance markers in pathogens worldwide. Understanding the dynamics of these microorganisms and their interaction with ticks is essential to effectively address the challenges related to veterinary and animal health. Additionally, it can provide valuable information for implementing more effective prevention and control strategies, thereby minimizing the impact of diseases transmitted by *R. sanguineus* s.l. on public health and the livestock industry. To gain a comprehensive understanding, it is imperative that future investigations focus on exploring the prevalence and circulation of the different pathogens identified in this study and implement active surveillance measures within the country.

Furthermore, conducting in-depth studies on the endosymbiont *C. mudrowiae* is crucial. This involves not only comprehending its behavior within the country through complete genome analysis but also assessing potential strategies that could pave the way for future vector control approaches utilizing this endosymbiont. Moreover, to further enhance our understanding of pathogen transmission in *R. sanguineus* s.l., it is imperative to expand this study to encompass

the diverse lineages of this tick species worldwide. This expansion would allow for the identification of distinct pathogens that may be transmitted by different lineages. Additionally, considering the knowledge generated regarding endosymbionts in these ticks, it becomes important to explore their prevalence and potential interactions within each lineage. By doing so, we can develop comprehensive surveillance strategies that encompass the entire *R. sanguineus* s.l. species complex, thereby enabling a more thorough assessment of pathogen diversity and transmission dynamics. Such an approach would provide valuable insights for the implementation of effective surveillance and control measures targeting this complex species.

Conclusion

Our study has yielded a comprehensive understanding of the pathogens and endosymbionts present in *R. sanguineus* s.l ticks in Colombia. We have successfully identified three significant pathogens, namely *A. phagocytophilum*, *F. tularensis*, and *T. equi*, which have implications for both animal and human health. Implications ranging from the occurrence of mild symptoms to the death of both humans and animals. This information is invaluable for guiding tick-borne disease surveillance efforts in the country. Furthermore, we have made notable discoveries regarding the presence of highly relevant endosymbionts, including *Coxiella mudrowiae*, among others, within this tick species. One intriguing finding is the negative correlations observed between the presence of pathogens and endosymbionts. This suggests complex interactions within the tick microbiome that warrant further investigation. In addition, our study represents a milestone as we have successfully assembled multiple MAGs for *C. mudrowiae*, making them the first MAGs of this endosymbiont reported in Colombia. These MAGs have provided valuable insights into the phylogenetic relationships and functional characteristics of *C. mudrowiae* in the country. Overall, our findings contribute significantly to the understanding of pathogen and endosymbiont dynamics in *R. sanguineus* s.l ticks in Colombia. This knowledge serves as a crucial resource for guiding tick-borne disease surveillance efforts and has implications for public health and animal welfare.

Acknowledgments

We thank the personnel who helped in the collection of the specimens in the field.

Additional information

Conflict of interest

The authors have declared that no competing interests exist.

Ethical statement

No ethical statement was reported.

Funding

This work was funded by Dirección de Investigación e Innovación from Universidad del Rosario.

Author contributions

Conceptualization: MM, JDR, LPT, APM. Data curation: MM, JDR, LPT. Formal analysis: JDR, LV, GH, LPT, MM. Funding acquisition: JDR. Investigation: JDR, GH, LV, APM, MM. Methodology: DGC, LPT, LV, APM, JDR, GH, MPM. Project administration: JDR. Resources: LPT, DGC, MPM, JDR. Software: LPT. Supervision: JDR, MM. Validation: JDR. Visualization: JDR, LPT, GH. Writing – original draft: JDR, LPT. Writing – review and editing: DGC, MPM, APM, GH, MM, LV.

Author ORCIDs

Diego Garcia-Corredor https://orcid.org/0000-0001-5122-5435
Juan David Ramírez https://orcid.org/0000-0002-1344-9312

Data availability

The raw sequences for each sample will be available under the project number PRJEB64029 in The European Nucleotide Archive (ENA) repository. The BioSampleID is specified in Suppl. material 6. MAGs will be available in a GitHub repository (https://github.com/luisa-p-triana/MAGs-Metagenomic-Tick).

References

Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen ALV, Cheng AA, Liu S, Min SY, Miroshnichenko A, Tran HK, Werfalli RE, Nasir JA, Oloni M, Speicher DJ, Florescu A, Singh B, Faltyn M, Hernandez-Koutoucheva A, Sharma AN, Bordeleau E, Pawlowski AC, Zubyk HL, Dooley D, Griffiths E, Maguire F, Winsor GL, Beiko RG, Brinkman FSL, Hsiao WWL, Domselaar GV, McArthur AG (2020) CARD 2020: Antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Research 48: D517–D525. https://doi.org/10.1093/nar/gkz935

Alneberg J, Bjarnason BS, De Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C (2014) Binning metagenomic contigs by coverage and composition. Nature Methods 11: 1144–1146. https://doi.org/10.1038/nmeth.3103

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of Molecular Biology 215(3): 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Andrews S (2010) FastQC: A Quality Control Tool for High Throughput Sequence Data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology 19(5): 455–477. https://doi.org/10.1089/cmb.2012.0021

Batool M, Blazier JC, Rogovska YV, Wang J, Liu S, Nebogatkin IV, Rogovskyy AS (2021) Metagenomic analysis of individually analyzed ticks from Eastern Europe demonstrates regional and sex-dependent differences in the microbiota of Ixodes ricinus. Ticks and Tick-Borne Diseases 12(5): e101768. https://doi.org/10.1016/j.ttbdis.2021.101768

Ben-Yosef M, Rot A, Mahagna M, Kapri E, Behar A, Gottlieb Y (2020) *Coxiella*-like endosymbiont of *Rhipicephalus sanguineus* is required for physiological processes during ontogeny. Frontiers in Microbiology 11: 1–16. https://doi.org/10.3389/fmicb.2020.00493

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30(15): 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bonilla-Aldana DK, Gutiérrez-Grajales EJ, Martínez-Arboleda JP, Reina-Mora MA, Trejos-Mendoza AE, Pérez-Vargas S, Valencia-Mejía L, Marín-Arboleda LF, Osorio-Navia D, Chacón-Peña M, González-Colonia LV, Cardona-Ospina JA, Jiménez-Posada EV, Diaz A, Salazar JC, Sierra M, Muñoz-Lara F, Zambrano LI, Ramírez-Vallejo E, Álvarez JC, Jaramillo-Delgado IL, Pecho-Silva S, Paniz-Mondolfi A, Faccini-Martínez ÁA, Rodríguez-Morales AJ (2022) Seroprevalence canine survey for selected vector-borne pathogens and its relationship with poverty in metropolitan Pereira, Colombia, 2020. Parasite Epidemiology and Control 17: 1–11. https://doi.org/10.1016/j.parepi.2022.e00249

Bonnet SI, Pollet T (2021) Update on the intricate tango between tick microbiomes and tick-borne pathogens. Parasite Immunology 43(5): e12813. https://doi.org/10.1111/pim.12813

Bonnet SI, Binetruy F, Hernández-Jarguín AM, Duron O (2017) The tick microbiome: Why non-pathogenic microorganisms matter in tick biology and pathogen transmission. Frontiers in Cellular and Infection Microbiology 7: 1–14. https://doi.org/10.3389/fcimb.2017.00236

Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA, Tringe SG, Ivanova NN, Copeland A, Clum A, Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P, Weinstock GM, Garrity GM, Dodsworth JA, Yooseph S, Sutton G, Glöckner FO, Gilbert JA, Nelson WC, Hallam SJ, Jungbluth SP, Ettema TJG, Tighe S, Konstantinidis KT, Liu WT, Baker BJ, Rattei T, Eisen JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, Lapidus A, Meyer F, Yilmaz P, Parks DH, Eren AM, Schriml L, Banfield JF, Hugenholtz P, Woyke T (2017) Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nature Biotechnology 35(8): 725–731. https://doi.org/10.1038/nbt.3893

Breitwieser FP, Salzberg SL (2020) Pavian: Interactive analysis of metagenomics data for microbiome studies and pathogen identification. Bioinformatics 36(4): 1303–1304. https://doi.org/10.1093/bioinformatics/btz715

Brenner AE, Muñoz-Leal S, Sachan M, Labruna MB, Raghavan R (2021) *Coxiella burnetii* and related tick endosymbionts evolved from pathogenic ancestors. Genome Biology and Evolution 13(7): 1–15. https://doi.org/10.1093/gbe/evab108

Buysse M, Duron O (2021) Evidence that microbes identified as tick-borne pathogens are nutritional endosymbionts. Cell 184(9): 2259–2260. https://doi.org/10.1016/j.cell.2021.03.053

Cabrera R, Mendoza W, López-Mosquera L, Cano MA, Ortiz N, Campo V, Keynan Y, López L, Rueda ZV, Gutiérrez LA (2022) Tick-borne-agents detection in patients with acute febrile syndrome and ticks from Magdalena Medio, Colombia. Pathogens 11(10): 1090. https://doi.org/10.3390/pathogens11101090

Cantalapiedra CP, Herņandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J (2021) eggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Molecular Biology and Evolution 38(12): 5825–5829. https://doi.org/10.1093/molbev/msab293

Cibichakravarthy B, Oses-Prieto JA, Ben-Yosef M, Burlingame AL, Karr TL, Gottlieb Y (2022) Comparative proteomics of *Coxiella* like endosymbionts (CLEs) in the symbiotic organs of *Rhipicephalus sanguineus* Ticks . Microbiology Spectrum 10: 1–11. https://doi.org/10.1128/spectrum.01673-21

Chao LL, Castillo CT, Shih CM (2021) Molecular detection and genetic identification of *Wolbachia* endosymbiont in *Rhipicephalus sanguineus* (Acari: Ixodidae) ticks of Taiwan. Experimental & Applied Acarology 83(1): 115–130. https://doi.org/10.1007/s10493-020-00574-3

Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH (2020) GTDB-Tk: A toolkit to classify genomes with the genome taxonomy database. Bioinformatics 36(6): 1925–1927. https://doi.org/10.1093/bioinformatics/btz848

Chen B, Jiang D-D, Liu Y-F, Lu X-Y, Yang G-P, Geng L, Wang X, Ren TG, Yang X (2023) High-throughput sequencing to detect the bacterial microbiome in Ixodes granulatus in China. Research Square: 1–15. https://doi.org/10.21203/rs.3.rs-2433236/v1

Cotes-Perdomo AP, Oviedo Á, Castro LR (2020) Molecular detection of pathogens in ticks associated with domestic animals from the Colombian Caribbean region. Experimental and Applied Acarology 82: 137–150. https://doi.org/10.1007/s10493-020-00531-0

Couper L, Swei A (2018) Tick microbiome characterization by next-generation 16S rRNA amplicon sequencing. Journal of Visualized Experiments 2018: 1–10. https://doi.org/10.3791/58239

Dantas-Torres F (2010) Biology and ecology of the brown dog tick, *Rhipicephalus sanguineus*. Parasites & Vectors 3(1): 1–3. https://doi.org/10.1186/1756-3305-3-26

Dantas-Torres F, Latrofa MS, Annoscia G, Giannelli A, Parisi A, Otranto D (2013) Morphological and genetic diversity of *Rhipicephalus sanguineus* sensu lato from the New and Old Worlds. Parasites & Vectors 6(1): e213. https://doi.org/10.1186/1756-3305-6-213

Ergunay K, Mutinda M, Bourke B, Justi SA, Caicedo-Quiroga L, Kamau J, Mutura S, Akunda IK, Cook E, Gakuya F, Omondi P, Murray S, Zimmerman D, Linton YM (2022) Metagenomic investigation of ticks from kenyan wildlife reveals diverse microbial pathogens and new country pathogen records. Frontiers in Microbiology 13: 1–16. https://doi.org/10.3389/fmicb.2022.932224

Ewels P, Magnusson M, Lundin S, Käller M (2016) MultiQC: Summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32(19): 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Gall CA, Reif KE, Scoles GA, Mason KL, Mousel M, Noh SM, Brayton KA (2016) The bacterial microbiome of *Dermacentor andersoni* ticks influences pathogen susceptibility. The ISME Journal 10(8): 1846–1855. https://doi.org/10.1038/ismej.2015.266

Ghafar MW, Amer SA (2012) Prevalence and first molecular characterization of *Anaplasma phagocytophilum*, the agent of human granulocytic anaplasmosis, in *Rhipicephalus sanguineus* ticks attached to dogs from Egypt. Journal of Advanced Research 3(2): 189–194. https://doi.org/10.1016/j.jare.2011.08.002

Gottlieb Y, Lalzar I, Klasson L (2015) Distinctive genome reduction rates revealed by genomic analyses of two *Coxiella*-like endosymbionts in ticks. Genome Biology and Evolution 7(6): 1779–1796. https://doi.org/10.1093/gbe/evv108

Gustavsen JA, Pai S, Isserlin R, Demchak B, Pico AR (2019) Rcy3: Network biology using cytoscape from within r. F1000 Research 8: 1–21. https://doi.org/10.12688/f1000research.20887.2 [version 1; peer review: 2 approved]

Hadfield J, Croucher NJ, Goater RJ, Abudahab K, Aanensen DM, Harris SR (2018) Phandango: An interactive viewer for bacterial population genomics. Bioinformatics 34(2): 292–293. https://doi.org/10.1093/bioinformatics/btx610

Herniou EA, Jehle JA (2007) Baculovirus Phylogeny and evolution. Curr Drug Targets 8: 1043–1050. https://doi.org/10.1007/0-306-47606-1_2

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution 23(2): 254–267. https://doi.org/10.1093/molbev/msj030

Hussain S, Perveen N, Hussain A, Song B, Aziz MU, Zeb J, Li J, George D, Cabezas-Cruz A, Sparagano O (2022) The symbiotic continuum within ticks: Opportunities for disease control. Frontiers in Microbiology 13: 1–17. https://doi.org/10.3389/fmicb.2022.854803

Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S (2018) High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nature Communications 9(1): e5114. https://doi.org/10.1038/s41467-018-07641-9

Jia N, Wang J, Shi W, Du L, Sun Y, Zhan W, Jiang JF, Wang Q, Zhang B, Ji P, Bell-Sakyi L, Cui XM, Yuan TT, Jiang BG, Yang WF, Lam TTY, Chang QC, Ding SJ, Wang XJ, Zhu JG, Ruan XD, Zhao L, Wei J, Te, Ye RZ, Que TC, Du CH, Zhou YH, Cheng JX, Dai PF, Guo W (2020) Large-scale comparative analyses of tick genomes elucidate their genetic diversity and vector capacities. Cell 182(5): 1328–1340.e13. https://doi.org/10.1016/j.cell.2020.07.023

Jolley KA, Bray JE, Maiden MCJ (2018) Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Research 3. https://doi.org/10.12688/wellcomeopenres.14826.1

Kanehisa M, Sato Y, Morishima K (2016) BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. Journal of Molecular Biology 428(4): 726–731. https://doi.org/10.1016/j.jmb.2015.11.006

Karim S, Budachetri K, Mukherjee N, Williams J, Kausar A, Hassan MJ, Adamson S, Dowd SE, Apanskevich D, Arijo A, Sindhu ZU, Kakar MA, Khan RMD, Ullah S, Sajid MS, Ali A, Iqbal Z (2017) A study of ticks and tick-borne livestock pathogens in Pakistan. PLoS Neglected Tropical Diseases 11: 1–17. https://doi.org/10.1371/journal.pntd.0005681

Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z (2019) MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ 2019: 1–13. https://doi.org/10.7717/peerj.7359

Karshima SN, Ahmed MI, Kogi CA, Iliya PS (2022) Anaplasma phagocytophilum infection rates in questing and host-attached ticks: A global systematic review and meta-analysis. Acta Tropica 228: e106299. https://doi.org/10.1016/j.actatropica.2021.106299

Kim D, Song L, Breitwieser FP, Salzberg SL (2016) Centrifuge: Rapid and sensitive classification of metagenomic sequences. Genome Research 26(12): 1721–1729. https://doi.org/10.1101/gr.210641.116

Lalzar I, Harrus S, Mumcuoglu KY, Gottlieb Y (2012) Composition and seasonal variation of *Rhipicephalus turanicus* and *Rhipicephalus sanguineus* bacterial communities. Applied and Environmental Microbiology 78: 4110–4116. https://doi.org/10.1128/AEM.00323-12

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nature Methods 9(4): 357–359. https://doi.org/10.1038/nmeth.1923

Letunic I, Bork P (2019) Interactive Tree of Life (iTOL) v4: Recent updates and new developments. Nucleic Acids Research 47(W1): W256–W259. https://doi.org/10.1093/nar/gkz239

Liu XY, Bonnet SI (2014) Hard tick factors implicated in pathogen transmission. PLoS Neglected Tropical Diseases 8(1): 1–5. https://doi.org/10.1371/journal.pntd.0002566

Liu B, Zheng D, Jin Q, Chen L, Yang J (2019) VFDB 2019: A comparative pathogenomic platform with an interactive web interface. Nucleic Acids Research 47(D1): D687–D692. https://doi.org/10.1093/nar/gky1080

Lopes de Carvalho I, Toledo A, Carvalho CL, Barandika JF, Respicio-Kingry LB, Garcia-Amil C, García-Pérez AL, Olmeda AS, Zé-Zé L, Petersen JM, Anda P, Núncio MS, Escudero R (2016) Francisella species in ticks and animals, Iberian Peninsula. Ticks and Tick-Borne Diseases 7(1): 159–165. https://doi.org/10.1016/j.ttbdis.2015.10.009

Luzzi M de C, de Carvalho LAL, Pinheiro DG, Lima-Duarte L, Camargo JV, Kishi LT, Fernandes CC, Machado RZ, Soares JF, André MR, Barros-Battesti DM (2021) Analysis on the prokaryotic microbiome in females and embryonic cell cultures of rhipicephalus sanguineus tropical and temperate lineages from two specific localities in brazil. Revista Brasileira de Parasitologia Veterinária 30(3): 1–15. https://doi.org/10.1590/s1984-29612021066

Machado-Ferreira E, Vizzoni VF, Balsemão-Pires E, Moerbeck L, Gazeta GS, Piesman J, Voloch CM, Soares CAG (2016) Coxiella symbionts are widespread into hard ticks. Parasitology Research 115(12): 4691–4699. https://doi.org/10.1007/s00436-016-5230-z

Máttar S, Parra M (2006) Detection of antibodies to *Anaplasma*, *Bartonella* and *Coxiella* in Rural Inhabitants of the Caribbean Area of Colombia. Revista Mvz Cordoba 11: 781–789. https://doi.org/10.21897/rmvz.441

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R (2020) IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Molecular Biology and Evolution 37: 1530–1534. https://doi.org/10.1093/molbev/msaa015

Molina-Guzmán LP, Ríos-Tobón S, Cardona-Lopera X, Lopera JA, Ríos-Osorio LA, Gutiérrez-Builes LA (2019) Occupational history of exposure to zoonotic agents in people dedicated to livestock in san pedro de los milagros, antioquia, Colombia. Revista de la Facultad de Medicina 67(4): 399–405. https://doi.org/10.15446/revfacmed.v67n4.72585

Nurk S, Meleshko D, Korobeynikov A, Pevzner PA (2017) MetaSPAdes: A new versatile metagenomic assembler. Genome Research 27(5): 824–834. https://doi.org/10.1101/gr.213959.116

Okonechnikov K, Golosova O, Fursov M (2012) Unipro UGENE: A unified bioinformatics toolkit. Bioinformatics 28(8): 1166–1167. https://doi.org/10.1093/bioinformatics/bts091

Olm MR, Brown CT, Brooks B, Banfield JF (2017) DRep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. The ISME Journal 11(12): 2864–2868. https://doi.org/10.1038/ismej.2017.126

Paez-Triana L, Muñoz M, Herrera G, Moreno-Perez D, Taffur-Gomez G, Montenegro D, Patarroyo MA, Paniz-Mondolfi A, Ramirez JD (2021) Genetic diversity and population structure of *Rhipicephalus sanguineus* s.l across different regions from Colombia. Parasites & Vectors 14(1): 1–11. https://doi.org/10.1186/s13071-021-04898-w

Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, Harris SR (2016) SNP-sites: Rapid efficient extraction of SNPs from multi-FASTA alignments. Microbial Genomics 2(4): e000056. https://doi.org/10.1099/mgen.0.000056

Papa A, Tsioka K, Daskou MA, Minti F, Papadopoulou E, Melidou A, Giadinis N (2020) Application of 16S rRNA next generation sequencing in ticks in Greece. Heliyon 6: e04542. https://doi.org/10.1016/j.heliyon.2020.e04542

Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW (2015) CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Research 25(7): 1043–1055. https://doi.org/10.1101/gr.186072.114

Pollet T, Sprong H, Lejal E, Krawczyk AI, Moutailler S, Cosson JF, Vayssier-Taussat M, Estrada-Peña A (2020) The scale affects our view on the identification and distribu-

tion of microbial communities in ticks. Parasites & Vectors 13(1): 1–13. https://doi.org/10.1186/s13071-020-3908-7

Portillo A, Palomar AM, de Toro M, Santibáñez S, Santibáñez P, Oteo JA (2019) Exploring the bacteriome in anthropophilic ticks: To investigate the vectors for diagnosis. PLoS ONE 14(3): e0213384. https://doi.org/10.1371/journal.pone.0213384

Posada D (2008) jModelTest: Phylogenetic model averaging. Molecular Biology and Evolution 25(7): 1253–1256. https://doi.org/10.1093/molbev/msn083

Rahal M, Medkour H, Diarra AZ, Bitam I, Parola P, Mediannikov O (2020) Molecular identification and evaluation of *Coxiella*-like endosymbionts genetic diversity carried by cattle ticks in Algeria. Ticks and Tick-Borne Diseases 11(5): e101493. https://doi.org/10.1016/j.ttbdis.2020.101493

Ravi A, Ereqat S, Al-Jawabreh A, Abdeen Z, Abu Shamma O, Hall H, Pallen MJ, Nasereddin A (2019) Metagenomic profiling of ticks: Identification of novel rickettsial genomes and detection of tick-borne canine parvovirus. PLoS Neglected Tropical Diseases 13(1): 1–19. https://doi.org/10.1371/journal.pntd.0006805

René-Martellet M, Minard G, Massot R, Van Tran V, Valiente Moro C, Chabanne L, Mavingui P (2017) Bacterial microbiota associated with *Rhipicephalus sanguineus* (s.l.) ticks from France, Senegal and Arizona. Parasites & Vectors 10(1): 1–10. https://doi.org/10.1186/s13071-017-2352-9

Rocafort-Ferrer G, Leblond A, Joulié A, René-Martellet M, Sandoz A, Poux V, Pradier S, Barry S, Vial L, Legrand L (2022) Molecular assessment of *Theileria equi* and *Babesia caballi* prevalence in horses and ticks on horses in southeastern France. Parasitology Research 121(3): 999–1008. https://doi.org/10.1007/s00436-022-07441-7

Rosenberg R, Lindsey NP, Fischer M, Gregory CJ, Hinckley AF, Mead PS, Paz-Bailey G, Waterman SH, Drexler NA, Kersh GJ, Hooks H, Partridge SK, Visser SN, Beard CB, Petersen LR (2018) Vital signs : Trends in reported vectorborne disease cases — United States and territories, 2004–2016. MMWR. Morbidity and Mortality Weekly Report 67(17): 496–501. https://doi.org/10.15585/mmwr.mm6717e1

Santodomingo A, Sierra-Orozco K, Cotes-Perdomo A, Castro LR (2019) Molecular detection of *Rickettsia* spp., *Anaplasma platys* and *Theileria equi* in ticks collected from horses in Tayrona National Park, Colombia. Experimental & Applied Acarology 77(3): 411–423. https://doi.org/10.1007/s10493-019-00354-8

Seemann T (2014) Prokka: Rapid prokaryotic genome annotation. Bioinformatics 30(14): 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Shannon P, Markiel A, Owen O, Baliga NS, Wang JT, Daniel R, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: A software environment for integrated models. Genome Research 13: 2498–504. https://doi.org/10.1101/gr.1239303.metabolite

Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF (2018) Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nature Microbiology 3(7): 836–843. https://doi.org/10.1038/s41564-018-0171-1

Stothard P, Grant JR, Van Domselaar G (2018) Visualizing and comparing circular genomes using the CGView family of tools. Briefings in Bioinformatics 20(4): 1576–1582. https://doi.org/10.1093/bib/bbx081

Telford SR III, Goethert HK (2020) Ecology of francisella tularensis. Annual Review of Entomology 65(1): 351–372. https://doi.org/10.1146/annurev-ento-011019-025134

Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, Gladstone RA, Lo S, Beaudoin C, Floto RA, Frost SDW, Corander J, Bentley SD, Parkhill J (2020) Producing polished prokaryotic pangenomes with the Panaroo pipeline. Genome Biology 21(1): 1–21. https://doi.org/10.1186/s13059-020-02090-4

Tsementzi D, Castro Gordillo J, Mahagna M, Gottlieb Y, Konstantinidis KT (2018) Comparison of closely related, uncultivated *Coxiella* tick endosymbiont population genomes reveals clues about the mechanisms of symbiosis. Environmental Microbiology 20(5): 1751–1764. https://doi.org/10.1111/1462-2920.14104

Venzal JM, Castro O, Cabrera PA, de Souza CG, Guglielmone AA (2003) Las garrapatas de Uruguay: Especies, hospedadores, distribución e importancia sanitaria. Veterinaria 30: 17–28.

Wade W (2002) Unculturable bacteria – The uncharacterized organisms that cause oral infections. Journal of the Royal Society of Medicine 95(2): 81–83. https://doi.org/10.1258/jrsm.95.2.81

Walker J, Keinrans J, Horak I (2000) The Genus *Rhipicephalus* (Acari, Ixodidae): A Guide to the Brown Ticks of the World. Cambridge University Press. https://doi.org/10.1017/CBO9780511661754

Wickham H (2008) Applied Spatial Data Analysis with R Elegant Graphics for Data Analysis: ggplot2: 21–54. https://doi.org/10.1007/978-0-387-98141-3

Wise LN, Kappmeyer LS, Mealey RH, Knowles DP (2013) Review of equine piroplasmosis. Journal of Veterinary Internal Medicine 27(6): 1334–1346. https://doi.org/10.1111/jvim.12168

Woolhouse MEJ, Gowtage-Sequeria S (2005) Host range and emerging and reemerging pathogens. Emerging Infectious Diseases 11(12): 1842–1847. https://doi.org/10.3201/eid1112.050997

Wu Y-W, Tang Y-H, Tringe SG, Simmons BA, Singer SW (2014) MaxBin: An automated binning method to recover individual genomes from metagenomes using. Microbiome 2(1): 4904–4909. https://doi.org/10.1186/2049-2618-2-26

Wu YW, Simmons BA, Singer SW (2016) MaxBin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32(4): 605–607. https://doi.org/10.1093/bioinformatics/btv638

Xia H, Hu C, Zhang D, Tang S, Zhang Z, Kou Z, Fan Z, Bente D, Zeng C, Li T (2015) Metagenomic Profile of the Viral Communities in *Rhipicephalus* spp. Ticks from Yunnan, China. PLoS ONE 10(3): e0121609. https://doi.org/10.1371/journal.pone.0121609

Zhang L, Liu H, Xu B, Lu Q, Li L, Chang L, Zhang X, Fan D, Li G, Jin Y, Cui F, Shi Y, Li W, Xu J, Yu XJ (2012) *Anaplasma phagocytophilum* infection in domestic animals in ten provinces/cities of China. The American Journal of Tropical Medicine and Hygiene 87(1): 185–189. https://doi.org/10.4269/ajtmh.2012.12-0005

Supplementary material 1

Geographic locations where samples were collected

Authors: Luisa Paez-Triana, Giovanny Herrera, Laura Vega, Diego Garcia-Corredor, Martin Orlando Pulido Medellín, Alberto Paniz-Mondolfi, Marina Muñoz, Juan David Ramirez

Data type: pdf

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.7.109085.suppl1

Supplementary material 2

Visual representation illustrating the methods utilized in the analysis discussed in the article

Authors: Luisa Paez-Triana, Giovanny Herrera, Laura Vega, Diego Garcia-Corredor, Martin Orlando Pulido Medellín, Alberto Paniz-Mondolfi, Marina Muñoz, Juan David Ramirez

Data type: pdf

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.7.109085.suppl2

Supplementary material 3

Significant differences in the relative abundances of assigned reads in Bacteria and Archaea between departments and sexes of the various genera found in the samples

Authors: Luisa Paez-Triana, Giovanny Herrera, Laura Vega, Diego Garcia-Corredor, Martin Orlando Pulido Medellín, Alberto Paniz-Mondolfi, Marina Muñoz, Juan David Ramirez

Data type: pdf

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.7.109085.suppl3

Supplementary material 4

Pathogens and endosymbionts present in each tick

Authors: Luisa Paez-Triana, Giovanny Herrera, Laura Vega, Diego Garcia-Corredor, Martin Orlando Pulido Medellín, Alberto Paniz-Mondolfi, Marina Muñoz, Juan David Ramirez

Data type: pdf

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.7.109085.suppl4

Supplementary material 5

Results of dereplication, as well as phylogenomics analysis and pangenome exploration of all samples (including longer MAGs)

Authors: Luisa Paez-Triana, Giovanny Herrera, Laura Vega, Diego Garcia-Corredor, Martin Orlando Pulido Medellín, Alberto Paniz-Mondolfi, Marina Muñoz, Juan David Ramirez

Data type: pdf

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.7.109085.suppl5

Supplementary material 6

Metadata and metagenomic information of collected samples

Authors: Luisa Paez-Triana, Giovanny Herrera, Laura Vega, Diego Garcia-Corredor, Martin Orlando Pulido Medellín, Alberto Paniz-Mondolfi, Marina Muñoz, Juan David Ramirez

Data type: xlsx

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.7.109085.suppl6

Supplementary material 7

Metadata and information regarding the genomes included in the analysis for endosymbiont and pathogen identification

Authors: Luisa Paez-Triana, Giovanny Herrera, Laura Vega, Diego Garcia-Corredor, Martin Orlando Pulido Medellín, Alberto Paniz-Mondolfi, Marina Muñoz, Juan David Ramirez

Data type: xlsx

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.7.109085.suppl7

Supplementary material 8

Abundance of taxonomy at the genus level within the domains of Bacteria and Archaea

Authors: Luisa Paez-Triana, Giovanny Herrera, Laura Vega, Diego Garcia-Corredor, Martin Orlando Pulido Medellín, Alberto Paniz-Mondolfi, Marina Muñoz, Juan David Ramirez

Data type: xlsx

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.7.109085.suppl8

Supplementary material 9

Abundance of taxonomy at the Family level within the domains of Bacteria and Archaea

Authors: Luisa Paez-Triana, Giovanny Herrera, Laura Vega, Diego Garcia-Corredor, Martin Orlando Pulido Medellín, Alberto Paniz-Mondolfi, Marina Muñoz, Juan David Ramirez

Data type: xlsx

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.7.109085.suppl9

Supplementary material 10

p-values obtained between departments and sexes within each department using the Wilcoxon test on the relative abundances of each genus (Bacteria and Archaea)

Authors: Luisa Paez-Triana, Giovanny Herrera, Laura Vega, Diego Garcia-Corredor, Martin Orlando Pulido Medellín, Alberto Paniz-Mondolfi, Marina Muñoz, Juan David Ramirez

Data type: xlsx

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.7.109085.suppl10

Supplementary material 11

Abundance of taxonomy of endosymbionts and pathogens

Authors: Luisa Paez-Triana, Giovanny Herrera, Laura Vega, Diego Garcia-Corredor, Martin Orlando Pulido Medellín, Alberto Paniz-Mondolfi, Marina Muñoz, Juan David Ramirez

Data type: xlsx

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.7.109085.suppl11

Supplementary material 12

p-values obtained between departments and sexes within each department using the Wilcoxon test on the relative abundances of pathogens and endosymbionts

Authors: Luisa Paez-Triana, Giovanny Herrera, Laura Vega, Diego Garcia-Corredor, Martin Orlando Pulido Medellín, Alberto Paniz-Mondolfi, Marina Muñoz, Juan David Ramirez

Data type: xlsx

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.7.109085.suppl12

Supplementary material 13

Relative abundance by sample of virulence factors

Authors: Luisa Paez-Triana, Giovanny Herrera, Laura Vega, Diego Garcia-Corredor, Martin Orlando Pulido Medellín, Alberto Paniz-Mondolfi, Marina Muñoz, Juan David Ramirez

Data type: xlsx

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.7.109085.suppl13

Supplementary material 14

Relative abundance by sample of resistence marker

Authors: Luisa Paez-Triana, Giovanny Herrera, Laura Vega, Diego Garcia-Corredor, Martin Orlando Pulido Medellín, Alberto Paniz-Mondolfi, Marina Muñoz, Juan David Ramirez

Data type: xlsx

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.7.109085.suppl14

Supplementary material 15

Characteristics of each of the metagenomes found and their comparison with available genomes

Authors: Luisa Paez-Triana, Giovanny Herrera, Laura Vega, Diego Garcia-Corredor, Martin Orlando Pulido Medellín, Alberto Paniz-Mondolfi, Marina Muñoz, Juan David Ramirez

Data type: xlsx

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.7.109085.suppl15

Supplementary material 16

Gene quantity analysis by COG clusters in MAGs, available genomes, and *C. burnetii*

Authors: Luisa Paez-Triana, Giovanny Herrera, Laura Vega, Diego Garcia-Corredor, Martin Orlando Pulido Medellín, Alberto Paniz-Mondolfi, Marina Muñoz, Juan David Ramirez

Data type: xlsx

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.7.109085.suppl16